AFRL-RH-WP-TR-2016-0093

AIR FORCE OFFICER QUALIFYING TEST FORM T:

INITIAL ITEM-, TEST-, FACTOR-, AND COMPOSITE-LEVEL ANALYSES

Thomas R. Carretta
Supervisory Control & Cognition Branch
Wright-Patterson AFB, OH

Mark R. Rose
John D. Trent
Air Force Personnel Center
Randolph AFB, TX

DECEMBER 2016
Interim Report

AIR FORCE RESEARCH LABORATORY
711 HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2016-0093 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

//signed//
ANTONIO AYALA
Work Unit Manager
Supervisory Control and Cognition Branch

//signed//
JASON B. CLARK
Chief, Supervisory Control and Cognition Branch
Warfighter Interface Division

//signed//
WILLIAM E. RUSSELL
Chief, Warfighter Interface Division
Airman Systems Directorate

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 28-11-2016	**2. REPORT TYPE** Interim	**3. DATES COVERED** *(From - To)* 8 July 2016 – 28 Nov 2016

4. TITLE AND SUBTITLE

Air Force Officer Qualifying Test Form T: Initial Item-, Test-, Factor-, and Composite-Level Analyses

5a. CONTRACT NUMBER FA8650-11-C-6158

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER 62202F

6. AUTHOR(S)

Thomas R. Carretta
Mark R. Rose
John D. Trent

5d. PROJECT NUMBER 5329

5e. TASK NUMBER 09

5f. WORK UNIT NUMBER H03K (53290902)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)

711 HPW/RHCI
2210 8th Street
Area B, Bldg. 146, Room 122
Wright-Patterson AFB, OH 45433-7511 (continued on next page)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Airman Systems Directorate
Warfighter Interface Division
Supervisory Control and Cognition Branch
Wright-Patterson AFB OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S) 711 HPW/RHCI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-RH-WP-TR-2016-0093

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

88ABW Cleared 01/23/2017; 88ABW-2017-0209. Report contains color.

14. ABSTRACT

The *Air Force Officer Qualifying Test* (AFOQT) is used to award scholarships to the United States Air Force (USAF) Reserve Officer Training Corps and to qualify applicants for officer commissioning through the ROTC and Officer Training School programs. The AFOQT also is used to qualify applicants for aircrew training as pilots, combat system operators, air battle managers, and remotely-piloted aircraft pilots, The purpose of this report is to document initial AFOQT Form T item-, test-, factor-, and composite level psychometric analyses. Data consisted of responses from USAF officer applicants who were administered either AFOQT Form T1 (N = 5,681) or Form T2 (N = 5,199) between 2015 and 2016. In general, both forms demonstrated acceptable psychometric properties. However, there were areas where improvements could be made. For example, item-level analyses revealed that the difficulty level for some (continued on next page)

15. SUBJECT TERMS

Air Force Officer Qualifying Test, AFOQT, psychometric analyses

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	SAR	48	Antonio Ayala
Unclassified	Unclassified	Unclassified			**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

7. Performing Organization(s) Names and Addresses:

Thomas R. Carretta
711 HPW/RHCI
2210 8th Street
Area B, Bldg. 146, Room 122
Wright-Patterson AFB, OH 45433-7511

Mark R. Rose
John D. Trent
Air Force Personnel Center
Strategic Research and Assessment Branch
AFPC/DSYX
Randolph AFB, TX 78150

14. Abstract (cont.):

Block Counting items was a function of an unfamiliar item presentation not used in the example items. Test- and composite-level analyses indicated that several scores had non-normal distribution shapes (skewness and kurtosis). For the most part this could be addressed by adding more difficult items to several tests. Results of the confirmatory factor analyses were consistent with previous forms. A model with five lower-order factors representing verbal, math, spatial, perceptual speed, and aviation knowledge, and a hierarchical general factor showed the best fit. However, some fit indices were below desired levels. This may have been due to the non-normality of the test score distributions and underrepresentation of the spatial factor.

CONTENTS

List of Figures

List of Tables

Preface

The analyses summarized in this report were conducted under work unit H03K (53290902) in support of enhanced airman alignment.

Air Force Officer Qualifying Test Form T:

Initial Item-, Test-, Factor-, and Composite-Level Analyses

The *Air Force Officer Qualifying Test* (AFOQT) is used to award scholarships to the US Air Force (USAF) Reserve Officer Training Corps (ROTC) and to qualify applicants for officer commissioning through the ROTC and Officer Training School (OTS) programs (United States Air Force, 2014). The AFOQT also is used to qualify applicants for aircrew training as pilots, combat system officers, air battle managers, and remotely-piloted aircraft pilots, if they pass other educational, fitness, medical, moral, and physical requirements. For operational use, the subtests are combined into six overlapping composites (see Table 1). The Verbal, Quantitative, and Academic Aptitude composites are used to qualify applicants for ROTC and OTS officer commissioning programs. The Pilot, Combat Systems Officer (CSO), and Air Battle Manager (ABM) composites are used to qualify applicants for aircrew training. The AFOQT has been validated against officer training performance (Roberts & Skinner, 1996), several aircrew training performance criteria including training completion (pass/fail), training grades, and class rank (Carretta, 2008, 2013; Carretta & Ree, 2003; Olea & Ree, 1994). It also has demonstrated predictive validity for several non-aviation officer jobs (Arth, 1986; Arth & Skinner, 1986; Carretta, 2010; Finegold & Rogers, 1985; Hartke & Short, 1988).

Since its implementation in 1953, the AFOQT has been revised several times, including numerous modifications to its content (see Drasgow, Nye, Carretta, & Ree, 2010). AFOQT Form T was implemented in 2015. The content of Form T differs from its immediate predecessor, Form S. Two spatial subtests that appeared on Form S, Rotated Blocks and Hidden Figures, were removed. Further, the General Science subtest (Form S) was modified to focus on the physical sciences (Physical Science) and a Reading Comprehension[1] subtest was added to improve assessment of verbal ability. Finally, the Situational Judgment[2] subtest was added to improve assessment of officership.

The purpose of this report is to document initial AFOQT Form T item-, test-, factor-, and composite level psychometric analyses. Item-level analyses included examination of item

[1] The Reading Comprehension subtest appeared on AFOQT Forma O, P, Q, and R. It was removed when Form S was implemented.
[2] The Situational Judgment subtest is experimental and is not included in this report.

difficulty, omission rate, and the item key and distractors. Test-level analyses included examination of score distribution shape, and internal consistency reliability. The factor-level analyses focused on evaluation of the latent factor structure of Form T and compared it to that of previous forms. Composite-level analyses focused on the distributional shape of the raw score composites with an eye toward determining whether the latent construct was adequately assessed throughout the ability range.

Table 1. AFOQT Composite Composition

		Composite					
Subtest	N Items	Pilot	CSO	ABM	Academic Aptitude	Verbal	Quant.
Verbal Analogies	25			X	X	X	
Arithmetic Reasoning	25				X		X
Word Knowledge	25		X		X	X	
Math Knowledge	25	X	X	X	X		X
Reading Comprehension	25				X	X	
Physical Science	20						
Table Reading	40	X	X	X			
Instrument Comprehension	25	X		X			
Block Counting	30		X	X			
Aviation Information	20	X		X			

Notes. Physical Science (PS) does not contribute to any of the AFOQT Form T composites. ABM = Air Battle Manager and CSO = Combat Systems Officer.

Methods

Participants

The data consisted of responses from US Air Force officer applicants who were administered either AFOQT Form T1 (*N* = 5,681) or Form T2 (*N* = 5,199) between 2015 and 2016. Scores were for those testing on the AFOQT for the first time. As summarized in Table 2 the demographic composition of the two samples was similar. The mean ages were 22.5 (T1) and 22.4 (T2) years and the mean education levels were 14.7 (T1) and 14.6 years (T2). All participants had completed at least 12 years of education. Both samples predominantly consisted of males (T1 = 75.2%; T2 = 75.3%) and Whites (T1 = 64.5%; T2 = 64.5%).

Table 2. Sample Demographic Data for AFOQT Forms T1 and T2

Variable	**Form T1 (*N* = 5,681)**		**Form T2 (*N* = 5,199)**	
	N	**%**	**N**	**%**
Sex				
Male	4,274	75.2	3,914	75.3
Female	1,399	24.6	1,279	24.6
Unknown	8	0.1	6	0.1
Race				
White	3,667	64.5	3,351	64.5
Black/African-American	710	12.5	686	13.2
Asian	570	10.0	522	10.0
Native-American/ Native-Alaskan	296	5.2	315	6.1
Native Hawaiian/ Other Pacific Islander	154	2.7	126	2.4
Unknown	284	5.00	199	3.83
Ethnicity				
Hispanic	747	13.1	693	13.3
Non-Hispanic	4,834	85.1	4.406	84.7
Unknown	100	1.8	100	1.9
Education				
Completed 12 Years (high school)	74	1.3	86	1.7
Completed 13 Years	1,830	32.2	1,698	32.7
Completed 14 Years	1,163	20.5	1,104	21.2
Completed 15 Years	604	10.6	544	10.5
Completed 16 Years	1,392	24.5	1.229	23.6
Completed 17 Years	322	5.7	268	5.2

Completed 18 Years	201	3.5	190	3.7
Completed 19 Years	41	0.7	37	0.7
Completed 20 Years	26	0.5	19	0.4
Completed 21+ Years	22	0.4	19	0.4
Unknown	6	0.1	5	0.1
Academic Degree				
High School Diploma	3,363	59.2	3,150	60.6
Associates Degree	438	7.7	376	7.2
Bachelor's Degree	1,659	29.2	1,464	28.2
Master's Degree	187	3.3	179	3.4
Unknown	16	0.2	17	0.3

Note. The percentages for Race do not add to 100% because respondents could choose more than one option and also could choose not to respond.

Measures

AFOQT Form T consists of 10 cognitive subtests that are combined into six operational composites (see Table 2). Personnel decisions including qualification for officer commissioning and aircrew training programs are based, in part, on AFOQT performance. Brief descriptions of the AFOQT subtests grouped by content are provided below.

Verbal Subtests

Verbal Analogies (VA) assesses the ability to reason and determine the relations between words. Word Knowledge (WK) measures verbal comprehension of written language involving the use of synonyms. Reading Comprehension (RC)[3] assesses the ability to read and understand written material.

Quantitative Subtests

Arithmetic Reasoning (AR) uses word problems to assess the ability to understand arithmetic relations. Math Knowledge (MK) assesses the ability to use mathematical formulas, relations, and terms.

[3] Reading Comprehension (RC) was an AFOQT subtest for Forms O through R. It was removed from Form S.

Spatial Subtest

Block Counting (BC) provides a measure of spatial ability through the analysis of three-dimensional representation of a set of blocks.

Aircrew Subtests

Instrument Comprehension (IC) measures the ability to determine the attitude of an aircraft from illustrations of flight instruments. Aviation Information (AI) assesses knowledge of general aviation concepts, principles, and terms. Physical Science (PS)[4] provides a measure of knowledge and understanding of scientific, terms, concepts, instruments, and principles.

Perceptual Speed Subtest

Table Reading (TR) measures the ability to quickly and accurately extract information from tables.

Analyses

Analyses were limited to first-time examinees. Item-level analyses began with an examination of item difficulty and omission rate. This was followed by examination of the item key and distractors. Test-level analyses focused on reliability of the scores, and shape of the score distributions. Internal consistency was examined for each subtest using Cronbach's alpha and item-total correlations. Test distribution shapes were assessed via examination of skewness and kurtosis.

Factor analyses examined the latent structure of the test. Several confirmatory factor models were examined and results were compared to those for previous forms. Composite-level analyses examined distributional shape (skewness and kurtosis) with a focus on whether the underlying aptitude was being assessed adequately across the aptitude range.

[4] General Science (GS), which appeared on Forms O through S, was revised with a focus on the physical sciences and was renamed Physical Science (PS).

Results

Item-Level Analyses

Item Difficulty and Item Omissions

Test items were scored as correct/incorrect (1/0). Items which were not answered (omissions) were scored as incorrect responses.

P-values. As summarized in Tables 3 and 4, p-values were similar for Forms T1 and T2. The most difficult subtests were AR, PS, BC, and AI. The higher difficulty for PS and AI are likely due to lack of prior exposure to their content. As discussed below, the higher difficulty for BC may be the result of speededness[5] of the subtest and item presentation (see Form Key and Distractors) for some items. For BC there are several items where the blocks touch on their back sides out of the participant's view. However, none of the example items illustrated this condition.

Item omissions. Item omissions were low for items 1-15 for each subtest, but generally increased throughout the subtest (see Tables 3 and 4). The subtests with the lowest omission rates were RC, PS, and AI. Those with the highest omission rates were TR, IC, and BC.

[5] Speededness is a test characteristic, dictated by a test's time limit, that results in a person's test score being dependent on the rate at which items are completed as well as the correctness of the responses.

Table 3. Subtest Item Difficulty Statistics: Form T1

	p values			Item Omissions (%)								
Subtest	Min,	Max.	Mean	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	Mean
VA	.30	.81	.604	0.42	0.50	0.76	2.58	7.10				2.272
AR	.27	.77	.558	0.50	1.10	1.40	2.71	5.94				2.330
WR	.35	.82	.608	0.40	0.42	0.54	1.42	2.82				1.120
MK	.25	.80	.586	0.76	0.53	1.02	1.25	3.04				1.320
RC	.37	.88	.687	0.11	0.09	0.07	0.11	0.45				0.167
PS	.33	.85	.550	0,07	0.18	0.12	0.29					0.165
TR	.16	.97	.676	0.12	0.15	0.34	1.56	5.29	13.98	24.35	33.15	9.867
IC	.33	.78	.605	0.20	0.29	2.08	7.53	16.23				5.267
BC	.08	.87	.513	0.22	0.09	0.89	4.30	12.65	23.25			6.900
AI	.30	.85	.477	0.16	0.16	0.29	1.00					0.322

N = 5,681

Table 4. Subtest Item Difficulty Statistics: Form T2

	p values			Item Omissions (%)								
Subtest	Min,	Max.	Mean	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	Mean
VA	.33	.81	.600	0.39	0.74	0.93	2.17	6.38				2.122
AR	.31	.76	.555	0,68	0.91	1.31	2.91	6.41				2.444
WR	.32	.85	.591	0.66	0.60	0.75	1.30	2.98				1.258
MK	.24	.81	.598	0.77	0.80	0.86	1.90	3.98				1.662
RC	.38	.93	.709	0.10	0.07	0.07	0.16	0.43				0.166
PS	.31	.86	.575	0.07	0.16	0.12	0.25					0.150
TR	.15	.97	.682	0.10	0.23	0.39	1.44	5.79	14.92	25.47	33.24	10.197
IC	.35	.79	.600	0.12	0.31	2.18	8.00	17.44				5.610
BC	.17	.91	.543	0.13	0.13	0.40	3.22	12.36	23.40			6.607
AI	.23	.77	.453	0.15	0.21	0.43	1.21					0.500

N = 5,199

Evaluation of Form Key and Distractors

There were several instances where one or more distractors was chosen more often than the keyed response. These generally occurred toward the end of the subtest where examinees may have been trying to complete the items before the time limit expired. A notable exception is BC, where the distinguishing feature of items with low accuracy was that the blocks touch on their back sides out of view of the participant. None of the example items illustrates this condition so some examinees may not have been aware that this could happen. See Table 5 for a summary.

Table 5. Summary of Form Key and Distractor Review

Subtest	Form T1		Form T2	
	Correct Choice Most Chosen	Notes	Correct Choice Most Chosen	Notes
Verbal Analogies	24 of 25	Item 24: one of the distractors was chosen more often than the keyed response	24 of 25	Item 23- one of the distractors was chosen more often than the keyed response
Arithmetic Reasoning	24 of 25	Items 23-25: one of the distractors was chosen almost as often as the keyed response	23 of 25	Items 23 & 24: one of the distractors was chosen more often than the keyed response
Word Knowledge	25 of 25		25 of 25	Items 23 & 24: One of the distractors was chosen nearly as often as the keyed response
Math Knowledge	24 of 25	An incorrect choice occurred nearly as often for items 22 and was chosen more often for item 24	24 of 25	Item 25: One of the distractors was chosen as often as the keyed response
Reading Comprehension	25 of 25		24 of 25	Item 20: one of the distractors

				was chosen more often than the keyed response
Physical Science	20 of 20		19 of 20	Item 18: A distractor was chosen more often than the keyed response. Also, one of the distractors was chosen frequently for items 19 and 20.
Table Reading	36 of 40	For 4 of the last 5 items, the most frequently chosen response “C” was not the correct response. Suggests patterned responding.	36 of 40	For 4 of the last 6 items, the most frequently chosen response “C” was not the correct response. Suggests patterned responding.
Instrument Comprehension	25 of 25		25 of 25	
Block Counting	24 of 30	One or more distractors chosen nearly as often (2) or more often (6) than the keyed choice.	23 of 30	One or more distractors chosen nearly as often (2) or more often (7) than the keyed choice.
Aviation Information	18 of 20	One of the distractors was chose as often or more often for items 17 and 18, and nearly as often as the keyed response for item 20.	17 of 20	There were 3 items for which the distractor was chosen more often than the keyed response (16, 17, & 19) and 2 items where the distractor was chosen nearly as often as the keyed response (9 & 18).

Subtest-Level Analyses

Descriptive Statistics

Table 6 summarizes the means, standard deviations (*SD*s), skewness, and kurtosis for the AFOQT Form T1 and T2 subtests. Examination of the skewness and kurtosis of the scores indicated that many of the distributions were non-normal, where the *t*-test for the skewness, kurtosis, or both exceeded +/- 1.96.

Table 7 shows the subtest correlations for each form. All correlations were positive. The subtest correlations had similar ranges and mean values for the two forms. The strongest correlations for both forms were between AR and MK (T1, $r = .705$; T2, $r = .748$) and the weakest were between WK and TR (T1, $r = .221$; T2, $r = .189$). The mean subtest correlations were .425 for Form T1 and .432 for Form T2. These values are very similar to those reported by Drasgow et al. (2010) for AFOQT Form S, where the correlations ranged from .706 (AR and MK) to .182 (WK and TR), with a mean of .413. These values also are similar to those for AFOQT Form Q which had 16 subtests, where the correlations ranged from .17 (WK and EM[6]) to .77 (WK and RC) with a mean of .436 (Carretta & Ree, 1996).

[6] EM is the Electrical Maze subtest. EM was removed from Form S.

Table 6. AFOQT Forms T1 and T2 Subtest Means, Standard Deviations, Skewness, and Kurtosis

Subtest	Form T1								Form T2							
	Mean	SD	Skew	Skew SE	Skew t-test	Kurt	Kurt SE	Kurt *t*-test	Mean	SD	Skew	Skew SE	Skew *t*-test	Kurt	Kurt SE	Kurt *t*-test
VA	15.11	4.42	-0.227	0.033	-6.88	-0.472	0.065	-7.26	14.88	4.63	-0.273	0.034	-8.02	-0.493	0.068	-7.25
AR	13.96	4.97	0.010	0.033	0.30	-0.681	0.065	-8.93	13.87	5.27	0.027	0.034	0.79	-0.747	0.068	-10.98
WK	15.22	5.55	-0.133	0.033	-4.03	-0.977	0.065	-15.03	14.76	5.55	-0.088	0.034	-2.58	-0.963	0.068	-14.16
MK	14.87	5.26	-0.060	0.033	-1.81	-0.892	0.063	-13.72	14.94	5.14	-0.081	0.034	-2.38	-0.814	0.068	-11.97
RC	17.18	4.00	-0.573	0.033	-17.36	0.000	0.065	0.00	17.74	4.30	-0.763	0.034	-22.44	0.287	0.068	4.22
PS	10.99	3.98	0.37	0.033	1.12	-0.835	0.065	-12.84	11.49	3.89	-0.058	0.034	-1.70	-0.723	0.068	-10.63
TR	27.05	5.86	-0.191	0.033	-5.78	-0.064	0.065	-0.98	27.26	6.50	-0.233	0.034	-6.85	0.036	0.068	0.53
IC	15.11	6.62	-0.361	0.033	-10.93	-0.917	0.065	-14.10	14.99	6.43	-0.257	0.034	-7.55	-1.022	0.068	-15.03
BC	15.38	5.79	-0.013	0.033	-0.39	-0.484	0.065	-7.13	16.29	5.27	-0.021	0.034	-1.00	-0.326	0.068	-4.79
AI	9.54	4.27	0.462	0.033	14.00	-0.557	0.065	-8.56	9.05	3.92	0.542	0.034	18.00	-0.231	0.068	-3.40

Notes. *t*-test values ≥+/-1.96 are statistically significant at $p \leq .05$
N T1 = 5,681; *N* T2 = 5,199

Table 7. AFOQT Forms T1 and T2 Subtest Correlations

Subtest	VA	AR	WK	MK	RC	PS	TR	IC	BC	AI
VA	1.000	0.530	0.715	0.506	0.651	0.515	0.289	0.390	0.382	0.349
AR	0.514	1.000	0.430	0.748	0.497	0.560	0.387	0.435	0.451	0.331
WK	0.670	0.437	1.000	0.393	0.654	0.485	0.189	0.339	0.300	0.340
MK	0.460	0.705	0.386	1.000	0.466	0.625	0.351	0.429	0.381	0.310
RC	0.605	0.475	0.645	0.410	1.000	0.487	0.268	0.402	0.329	0.367
PS	0.503	0.539	0.485	0.620	0.475	1.000	0.208	0.474	0.334	0.454
TR	0.317	0.459	0.221	0.396	0.283	0.263	1.000	0.340	0.439	0.229
IC	0.393	0.450	0.295	0.407	0.379	0.488	0.478	1.000	0.492	0.572
BC	0.354	0.428	0.265	0.351	0.298	0.298	0.517	0.504	1.000	0.311
AI	0.346	0.342	0.343	0.302	0.411	0.444	0.304	0.560	0.307	1.000

Note. The Form T1 subtest correlations are below the diagonal and the Form T2 subtest correlations are above the diagonal.

N T1 = 5.681; *N* T2 = 5,199

Internal Consistency

Internal consistency results were similar for Forms T1 and T2 (see Tables 8 and 9). Cronbach's alpha ranged from .730 (RC) to .913 (IC) for Form T1 and from .741 (AI) to .904 (IC) for Form T2 with respective mean reliabilities of .816 and .815. Six subtests (AR, WK, MK, TR, IC, and BC) had reliabilities of .80 or higher for both forms.

The lowest item-total correlations for Form T1 occurred for VA (.367), RC (.379), and TR (.354) and the highest occurred for WK (.473) and IC (.570). The lowest item-total correlations for Form T2 were for VA (.369). RC (.403), and BC (.399); the highest were for MK (.452) and IC (.552).

Table 8. Subtest Internal Consistency: Form T1

Subtest	N Items	Cronbach's	Item-Total Correlations		
		Alpha	Min.	Max,	Mean
VA	25	.740	.214	.459	.367
AR	25	.804	.312	.504	.421
WK	25	.856	.327	.592	.473
MK	25	.838	.304	.574	.455
RC	25	.730	.209	.482	.359
PS	20	.759	.225	.606	.423
TR	40	.883	.115	.673	.374
IC	25	.913	.318	.653	.570
BC	30	.847	.282	.532	.428
AI	20	.790	.292	.613	.447

N = 5.681

Table 9. Subtest Internal Consistence: Form T2

Subtest	N Items	Cronbach's	Item-Total Correlations		
		Alpha	Min.	Max.	Mean
VA	25	.769	.284	.542	.369
AR	25	.830	.304	.573	.420
WK	25	.851	.310	.608	.468
MK	25	.827	.298	.623	.452
RC	25	.781	.233	.527	.403
PS	20	.745	.282	.554	.413
TR	40	.887	.201	.678	.414
IC	25	.904	.309	.661	.552
BC	30	.822	.191	.597	.399
AI	20	.741	.270	.555	.411

N = 5,199

Latent Factor Structure of Form T and Comparison with Previous Forms

Skinner and Ree (1987) conducted an exploratory factor analysis of Form O on a sample of 3,000 US Air Force officer commissioning applicants. They reported a five-factor solution for the 16 Form O subtests: verbal, math, spatial, aircrew interests/aptitude, and perceptual speed. Correlations between the factors ranged from .22 to .50, with a mean of .36. Noting the correlations among the factors, Carretta and Ree (1996) reanalyzed the Skinner and Ree (1987) data using confirmatory factor analysis methods. Several models were specified and estimated. They included a single factor model (psychometric *g*), a four-factor model reflecting the AFOQT operational composites (Verbal, Quantitative, Pilot, and Navigator/Technical), a five-factor model of verbal, math, spatial, aircrew interests/aptitude, and perceptual speed (Skinner & Ree, 1987), a bifactor model with the four operational composites and *g*, and a bifactor model with the five Skinner and Ree factors and *g*. The model with *g* and five content factors (verbal, math, spatial, aviation, and perceptual speed) provided a good fit to the data with a root mean square error of approximation (RMSEA) of .071, a comparative fit index (CFI) of .957, and an average absolute standardized residual of .027.

When AFOQT Form S was implemented, five subtests that appeared on Forms O through R had been removed (Reading Comprehension, Data Interpretation, Mechanical Comprehension, Electrical Maze, and Scale Reading) to shorten test administration. As discussed by Drasgow et al. (2010), confirmatory factor analyses (CFAs) of AFOQT Form S presented a challenge because two of its content factors (math and perceptual speed) were expected to have nonzero loadings for only two subtests, whereas at least three nonzero loadings are needed for statistical estimation of factor loadings. Drasgow et al. used exhaustive and mutually-exclusive sets of items to create multi-item composites (called “item parcels” by Dorans & Lawrence, 1987) for each subtest. These multi-item composites (parcels) were then factor-analyzed. For example, five parcels were created for Word Knowledge (25 items) and eight parcels for Table Reading (40 items). Because there were five parcels each for the Arithmetic Reasoning and Math Knowledge subtests, factor loadings could be estimated for 10 scores for the mathematical reasoning factor. As a result, the factor loadings were statistically identified. Drasgow et al. evaluated several CFAs and concluded that the data were best represented by a bifactor model with a general factor and five content factors representing verbal, math, spatial, aircrew, and perceptual speed (RMSEA = .053, CFI = .98, and SRMR = .057).

The problem of too few subtests to adequately specify some content factors also occurred for AFOQT Form T where nonzero loadings were expected for only two subtests for the math and perceptual speed factors, and one subtest for the spatial[7] factor. In order to examine the latent structure of Forms T1 and T2, we followed the approach used by Drasgow et al. (2010) of analyzing multi-item composites in lieu of subtest scores. As with Form S, the large number of items (N = 260) precluded factor analyses using item-level data. Exhaustive and mutually-exclusive sets of items were used to create multi-item composites (item parcels) for each subtest which were then factor-analyzed.

Procedures

Several confirmatory factor analysis (CFAs) were examined to evaluate the structure of AFOQT Forms T1 and T2. The starting model consisted of a factor representing general cognitive ability (*g*) and five specific cognitive factors of verbal, math, spatial, aircrew knowledge, and perceptual speed. This model was based on a confirmatory model of the previous 16 subtest version (Carretta & Ree, 1996) and 11 subtest version (Drasgow et al., 2010) of the AFOQT. Based on CFA results for AFOQT Form S (Drasgow et al., 2010), BC and PS were allowed to cross-load on more than one lower-order factor. The lower-order factors were defined as: verbal (VA, WK, RC, and PS), math (AR and MK), spatial (BC), aviation (PS, IC, and AI), and perceptual speed (TR and BC).

Analyses

Analyses began with an examination of the subtest correlations for each form. The *g*-saturation of the forms was estimated from the first unrotated principal component as discussed by Ree and Earles (1991).

Several CFAs were examined. Model 1 had a single general factor (*g*) on which all 52 item parcels directly loaded. Model 2 consisted of four content factors representing verbal,

---

[7] AFOQT Form S had three spatial subtests – Block Counting (BC), Rotated Blocks (RB), and Hidden Figures (HF). AFOQT Form T has only one spatial subtest, BC.

math, aviation, and perceptual speed. It was tested because whereas AFOQT Form S had three spatial subtests (BC, RB, and GS), Form T has only one (BC). Model 3 consisted of five content factors (verbal, math, spatial, aviation, and perceptual speed) which is consistent with previous AFOQT forms (Carretta & Ree, 1996; Drasgow et al., 2010; Skinner & Ree, 1987). Model 4 was Model 2 (4 content factors) with a hierarchical general factor derived from the lower-order factors. Model 5 was Model 3 (5 content factors) with a hierarchical general factor derived from the lower-order factors. The examination of models with a hierarchical general factor differs from Carretta and Ree (1996) and Drasgow et al. (2010) who employed a bifactor model, where the test scores loaded on both a general factor and specific factor.

The models were estimated using maximum likelihood. Two important issues for structural equation modeling are the degree to which the models are correctly specified and the data are multivariate normal. Maximum Likelihood (ML) and Generalized Least Squares (GLS) estimation procedures will produce similar results when the hypothesized model is correctly specified and the observed variables are multivariate normal (Olsson, Foss, Troye, & Howell, 2000). When these conditions are not met ML and GLS may not converge on the same optimal solution. In a simulation study, Olsson et al. examined the effect of estimation method on parameter estimation and model fit for varying sample sizes, amount of specification error, and level of kurtosis. They concluded that under conditions of misspecification, ML compared with GLS provides more realistic indices of overall fit and less biased parameter values for paths that overlap with the true model. Olsson et al. further stated that despite recommendations in the literature that weighted lease squares (WLS) estimation be used when data are not normally distributed, under no conditions was it preferable to ML or GLS in regard to parameter bias and fit.

Several goodness-of-fit statistics were examined to evaluate model fit. The choice of indices was guided, in part, by Hu and Bentler (1998, 1999) who recommend using both an absolute fit index and an incremental fit index to examine model fit. We chose the absolute fit indices of the Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1993), Standardized Root Mean Square Residual (SRMR; Hu & Bentler, 1999), Critical N (Jöreskog & Sörbom, 1989), Goodness-of-Fit Index (GFI; Tanaka & Huba, 1985), and Adjusted Goodness-of-fit Index (AGFI; Jöreskog & Sörbom, 1989). The incremental fit indices chosen were the

Comparative Fit Index (CFI; Bentler, 1990, 1995) and Non-Normed Fit Index (NNFI; Bentler, 1990, 1995). The RMSEA, CFI, GFI, AGFI, NNFI, and Critical N fit indices are sensitive to misspecification of factor loadings. The SRMR is sensitive to misspecification of factor covariances (Hu & Bentler, 1998). Hu and Bentler (1999) recommended the following cutoff values as indicators of good model fit: CFI and NNFI ≥ .95, SRMR ≤ .08, and RMSEA ≤ .06. In addition, previous research suggested that a GFI ≥ .95 (Marsh & Grayson, 1995) and an AGFI ≥ .90 (Schermelleh-Engel, Moosbrugger, & Muller, 2003) indicate acceptable model fit.

Results: Descriptive Statistics

As previously discussed and summarized in Table 6, values for skewness and kurtosis of the subtest scores indicated that many of the distributions were non-normal, where the *t*-test for the skewness, kurtosis, or both exceeded +/- 1.96. Similar results were obtained for the scores based on item parcels.

Table 7 presents the subtest correlations for each form. The subtest correlation matrix for the item parcels is available from the first author. All correlations were positive. As previously noted, similar ranges and mean values were observed for the subtest correlations for the two forms. The strongest correlations for both forms were between AR and MK (T1, $r = .705$; T2, $r = .748$). The weakest correlations for both forms were between WK and TR (T1, $r = .221$; T2, $r = .189$). The mean subtest correlations were .425 for Form T1 and .432 for Form T2. These values are very similar to those reported for AFOQT Form S (Drasgow et al., 2010), where the correlations ranged from .706 (AR and MK) to .182 (WK and TR), with a mean of .413. These values also are similar to those for AFOQT Form Q which had 16 subtests. The subtest correlations for Form Q ranged from .17 (WK and EM) to .77 (WK and RC) with a mean of .436 (Carretta & Ree, 1996).

Results: g-saturation

The *g*-saturation of AFOQT Forms T1 and T2 was estimated from the first unrotated principal component as discussed by Ree and Earles (1991). An eigenvalue analysis of the subtest correlations indicated that general cognitive ability (*g*) accounted for 48.2% of the variance for Form T1 and 48.4% for Form T2. These results were very similar to the value of 47% reported for Form S (Drasgow et al., 2010). An examination of the communalities indicated that the highest values occurred for the three verbal subtests, WK (T1 = .728, T2 = .738), VA (T1 = .685, T2 = .732), and RC (T1 = .668, T2 = .682) and the lowest for AI (T1 = .385, T2 = .395).

The *g*-saturation also was estimated for the 52 item parcels since these were the scores used in the CFAs. The percent of variance accounted for the first unrotated factor was 28.1% for Form T1 and 26.5% for Form T2. Drasgow et al. (2010) did not report the percent of variance accounted for by *g* for their item parcels.

Results: Confirmatory Factor Analyses

Tables 10 and 11 summarize the fit statistics for Forms T1 and T2. Model fit for Form T1 was somewhat poorer than for Form T2. The reason for this is unknown, but may be due to sample composition.

The single factor model fit the data poorly for both forms. RMSEA values of .095 (T1) and .094 (T2) and SRMR values of .099 (T1) and .099 (T2) were above the values for a good fit recommended by Hu and Bentler (1999). The values for the other indices (CFI, GFI, AGFI, NNFI, and Critical n) were well below recommended values for a good fit.

Table 10. Fit Statistics for AFOQT Form T1 CFAs using Item Parcels for Maximum Likelihood (ML) Estimation

Model	RMSEA	CFI	GFI	AGFI	NNFI	SRMR	Critical N
M1: *g*	0.095	0.54	0.51	0.47	0.52	0.099	120.51
M2: 4 lower-order factors	0.061	0.81	0.77	0.75	0.80	0.072	279.88
M3: 5 lower-order factors	0.050	0.87	0.87	0.86	0.86	0.067	406.64
M4: *g* + 4 lower-order factors	0.061	0.81	0.77	0.78	0.80	0.072	277.59
M5: *g* + 5 lower-order factors	0.050	0.87	0.87	0.86	0.86	0.067	404.53

Model fit was best for Models 3 (5 lower-order factors) and 5 (5 lower-order factors with a hierarchical factor). Fit statistics for these models were in the acceptable range for both forms for the RMSEA, SRMR, and Critical N. However, the CFI, GFI, AGFI, and NNFI were below recommended values for both forms.

As previously discussed, the skewness and kurtosis values for the AFOQT subtests and parcels indicated that the distributions for several of the scores were non-normal. ML estimation is not optimal under this condition.

Table 11. Fit Statistics for AFOQT Form T2 CFAs using Item Parcels for Maximum Likelihood (ML) Estimation

Model	RMSEA	CFI	GFI	AGFI	NNFI	SRMR	Critical N
M1: *g*	0.094	0.55	0.51	0.47	0.53	0.099	121.47
M2: 4 lower-order factors	0.055	0.85	0.82	0.80	0.84	0.067	346.54
M3: 5 lower-order factors	0.043	0.91	0.89	0.88	0.90	0.045	535.21
M4: *g* + 4 lower-order factors	0.055	0.85	0.82	0.80	0.84	0.069	344.81
M5: *g* + 5 lower-order factors	0.043	0.91	0.89	0.88	0.90	0.046	532.91

Discussion: Confirmatory Factor Analyses

Analyses began with an examination of the subtest correlations and an eigenvalue analysis of AFOQT Forms T1 and T2. Next, several confirmatory factor analytic models were fit to the data. In general, the results were consistent with those for earlier AFOQT forms (Carretta & Ree, 1996; Drasgow et al., 2010). The range and mean value of the subtest correlations and the *g*-saturation for Forms T1 and T2 were very similar to previous forms. Results supported the existence of a general cognitive ability factor that underlies all of the subtests and verbal, math, spatial, aircrew, and perceptual speed factors that underlie groups of subtests. However, model fit, especially for Form T1, was not as good as observed for earlier forms.

The reasons for the somewhat lower fit are not clear, but may be due to changes in content from previous forms and/or non-normality of the subtest score distributions. Although Form T shares several subtests (VA, AR, WK, MK, TR, IC, BC, and AI) with Form S, two of the spatial subtests (Rotated Blocks and Hidden Figures) that appeared on Form S were dropped from Form T and General Science was modified to focus on physical sciences (PS). Model fit may have been adversely affected with fewer indicators of spatial ability and modified science content. Model fit also may have been adversely affected by extreme skewness and kurtosis values for several of the subtests. Olsson et al. (2000) found ML estimation to be robust in parameter estimation and model fit under varying levels of misspecification and kurtosis in a simulation study. However, Olsson et al. did not examine the joint effects of extreme skewness and kurtosis as occurred with many of the Form T scores. In a Monte Carlo simulation, Benson and Fleishman (1994) found that under conditions of non-normality (increases in skewness and kurtosis), standard error was underestimated and ML chi-square statistics were inflated.

Another explanation for somewhat poorer model fit for Form T compared with Form S (Drasgow et al., 2010) may be due to the way the item parcels were constructed in the two studies. Drasgow et al. created parcels by grouping consecutive items in sets of 5 such as items 1-5, 6-10, 11-15, 16-20, and 20-25. In the current study item parcels were created where the items came from different parts of the test (e.g. parcel 1 consisted of items 1, 7, 12, 17, and 22). An examination of item-level data revealed that the rate of item omission and guessing increases for later items. Examinees may be running out of time so either skip items or guess towards the later items. Thus, when Drasgow et al, created their item parcels, the early parcels are likely more reliable than the later parcels where guessing and item omission are more prevalent. In contrast, since the parcels in the current study sample items throughout the test, these less reliable items are dispersed across the parcels. This sampling may have in turn affected the CFA parameter estimates and model fit.

Raw Score Composite-Level Analyses

Examination of the AFOQT composites revealed that with the exception of the Quantitative composite all were significantly negatively skewed. All six composites showed some truncation at the upper end of their distribution (ceiling effect), though this was greatest for the Verbal and Quantitative composites. Results for Form T1 are summarized in Table 12 and Figures 1-6; those for Form T2 are summarized in Table 13 and Figures 7-12. The figures show a normal curve imposed over the score distributions.

On the subtest level, skewness was largest for RC (-) and AI (+) for both Forms T1 and T2. Several subtests showed large effects for kurtosis. The largest effects for both forms occurred for WK (-), MK (-), PS (-), and IC (-). In general, the distribution shapes for the AFOQT composites could be improved if additional more difficult items were added. Improving distributional shape and discriminability is more important for the aviation-related composites (Pilot, CSO, and ABM) than for the Verbal. Quantitative, or Academic Aptitude composites.

Table 12. Shape of AFOQT T1 Raw Score Composite Distributions

	Composite					
Statistic	Verbal	Quantitative	Academic	Pilot	CSO	ABM
Mean	15.84	14.57	15.20	16.38	18.30	17.88
St. Dev.	4.07	4.82	3.90	4.17	4.13	3.97
Skewness	-0.267	-0.016	-0.153	-0.141	-0.157	-0.189
SE Skewness	0.033	0.033	0.033	0.033	0.033	0.033
Skewness *t*	-8.09***	-0.48	-4.64***	-4.27***	-4.76***	-5.73***
Kurtosis	-0.656	-0.833	-0.647	-0.654	-0.329	-0.436
SE Kurtosis	0.065	0.065	0.065	0.065	0.065	0.065
Kurtosis *t*	-10.09***	-12.81***	-9.95***	-10.06	-5.06***	-6.71***

N = 5,691; ***$p \leq .001$

Table 13. Shape of AFOQT T2 Raw Score Composite Distributions

	Composite					
Statistic	Verbal	Quantitative	Academic	Pilot	CSO	ABM
Mean	15.80	14.58	15.19	16.26	18.48	17.92
St. Dev.	4.28	4.88	4.04	4.03	4.07	3.94
Skewness	-0.338	-0.006	-0.184	-0.130	-0.158	-0.204
SE Skewness	0.034	0.034	0.034	0.034	0.034	0.034
Skewness *t*	-9.94***	-0.17	-5.41***	-3.82***	-4.65***	-6.00***
Kurtosis	-0.572	-0.835	-0.635	-0.597	-0.291	-0.390
SE Kurtosis	0.068	0.068	0.068	0.068	0.068	0.068
Kurtosis *t*	-15.05***	-12.72***	-9.34***	-8.78***	-4.28***	-5.73***

N = 5,199; ***$p \leq .001$

The minimum qualifying scores for officer commissioning programs are relatively low for the Verbal (15) and Quantitative (10) composites. Discriminability is most important around the minimum qualifying score. Adding difficult verbal or math items would not improve

discriminability for these composites. In contrast, competition for aircrew training assignments is much stronger. Although minimum qualifying scores are relatively low for aircrew training (e.g., Pilot ≥ 25), in practice the mean Pilot composite score for those accepted into pilot training is about 80. Therefore, it is more important to improve discriminability for the aviation-related composites in the high aptitude range (i.e., greater than 70). To do so, additional difficult items are needed for some subtests that contribute to the aircrew-related composites (Pilot, CSO, and ABM).

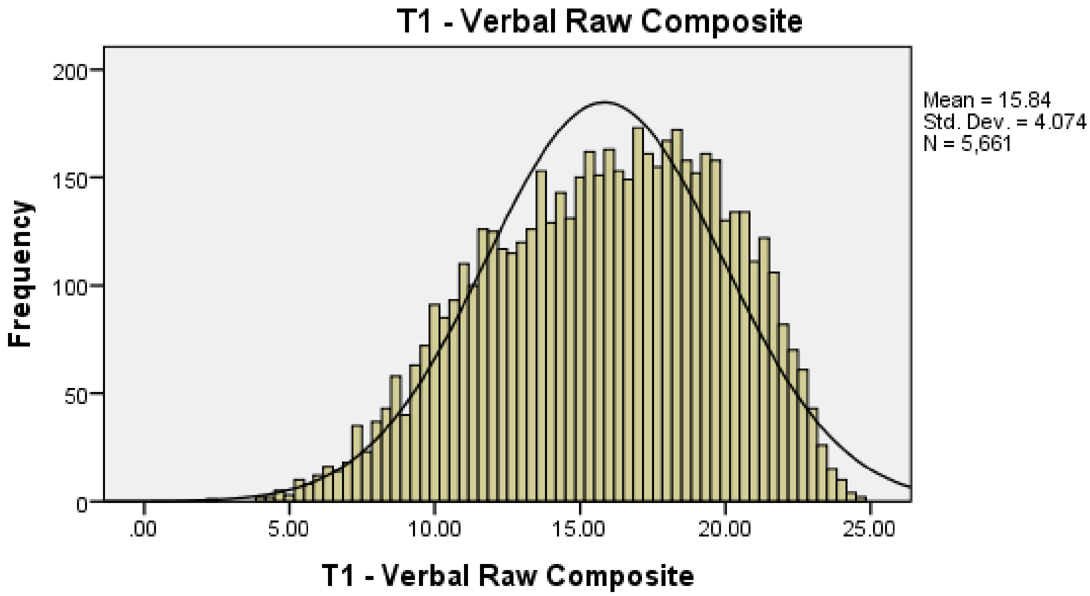

Figure 1. AFOQT Form T1 Verbal raw composite score distribution.

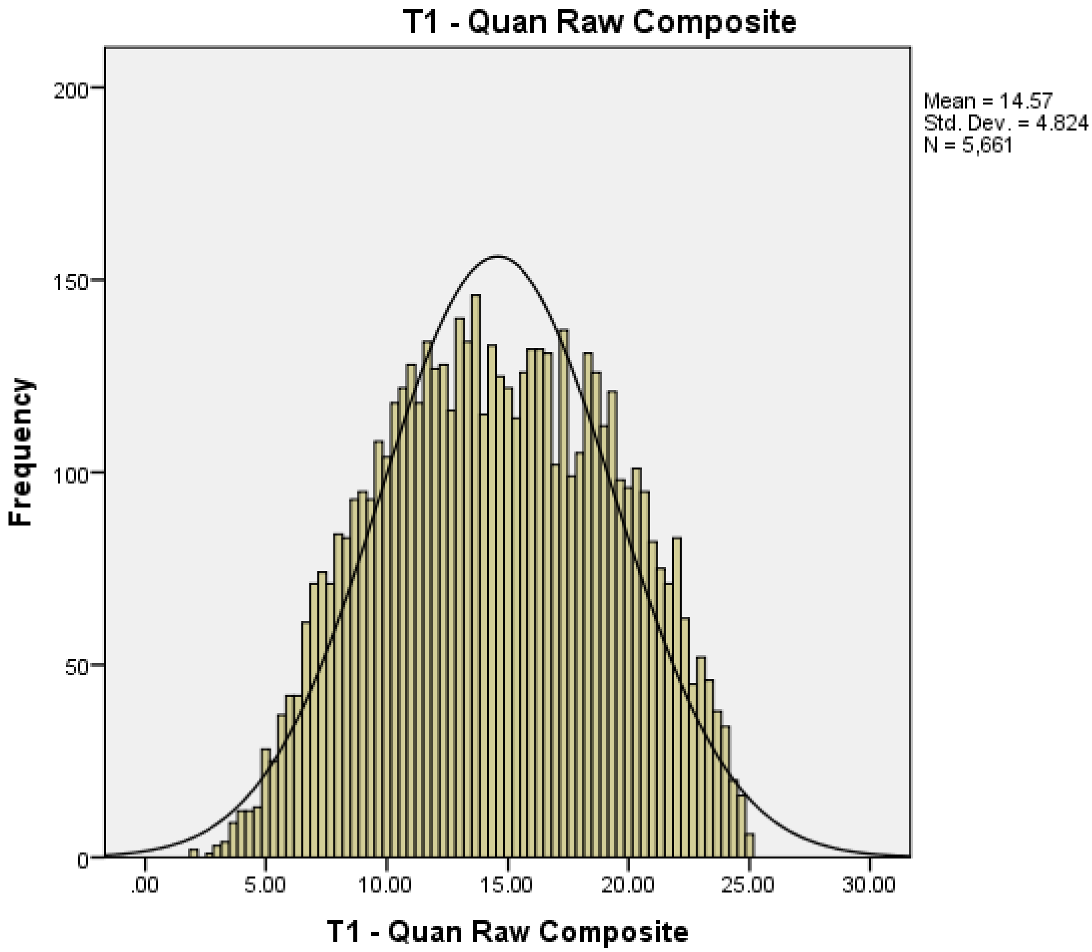

Figure 2. AFOQT Form T1 Quantitative raw composite score distribution.

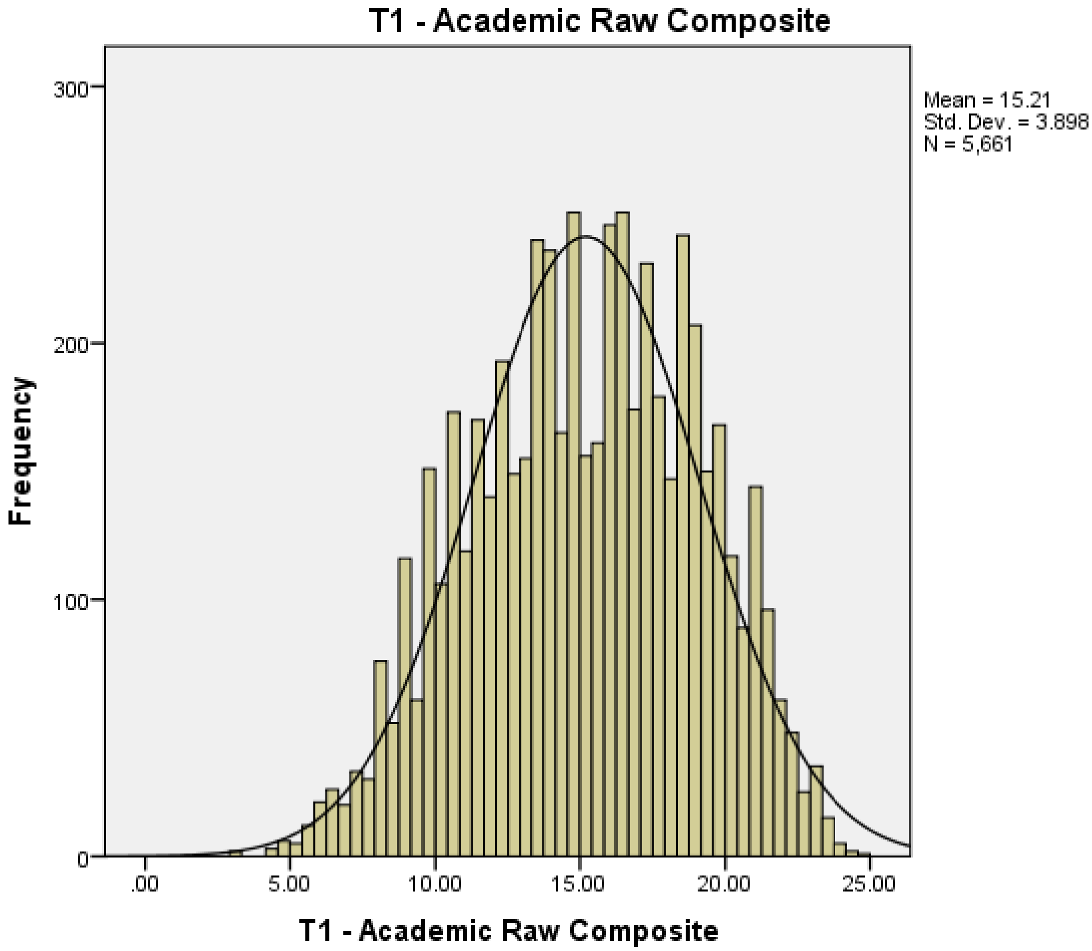

Figure 3. AFOQT Form T1 Academic Aptitude raw composite score distribution.

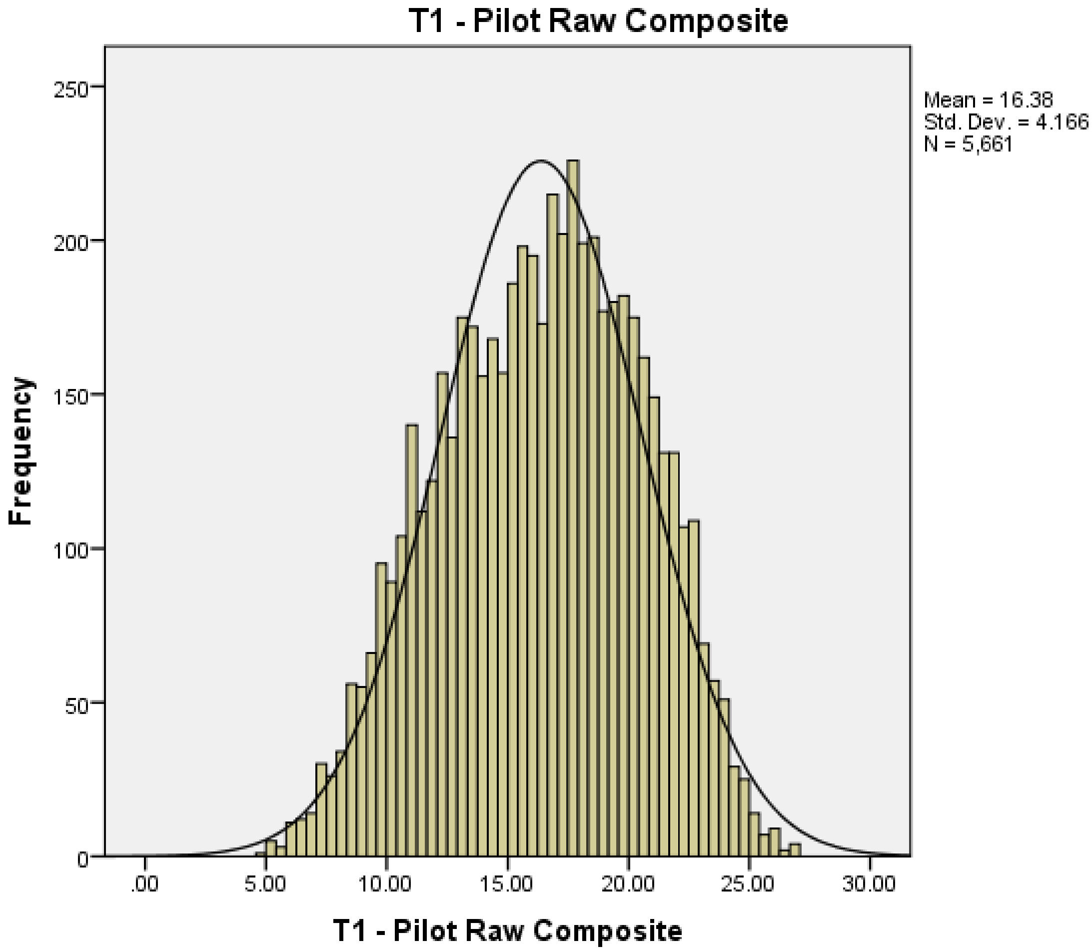

Figure 4. AFOQT Form T1 Pilot raw composite score distribution.

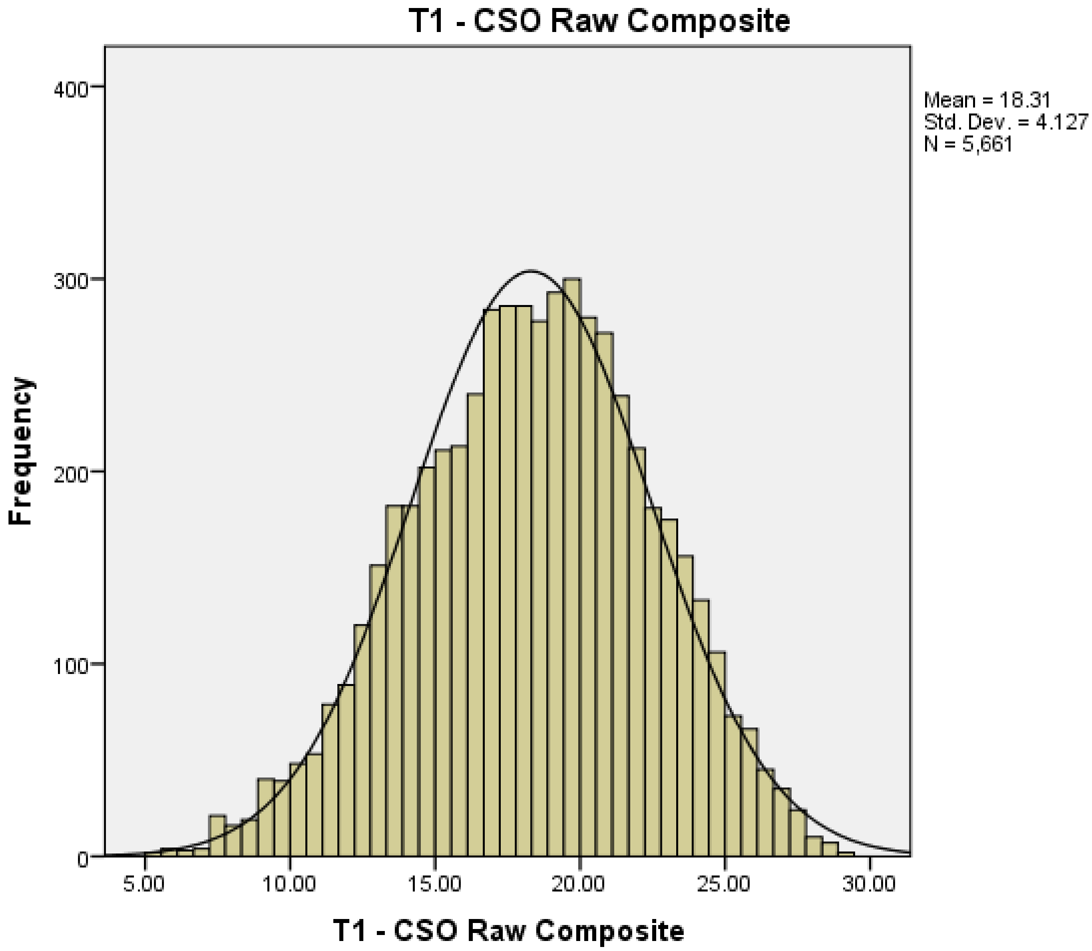

Figure 5. AFOQT Form T1 Combat Systems Officer (CSO) raw composite score distribution.

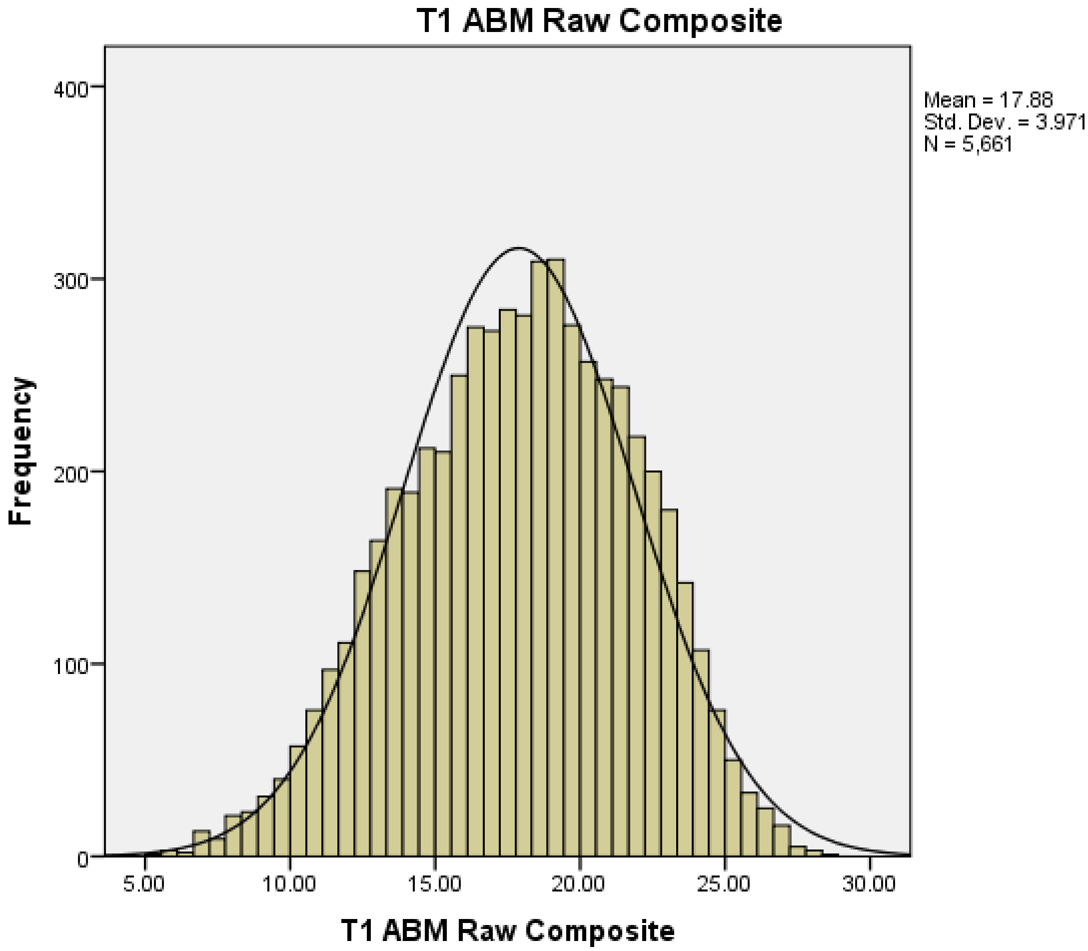

Figure 6. AFOQT Form T1 Air Battle Manager (ABM) raw composite score distribution.

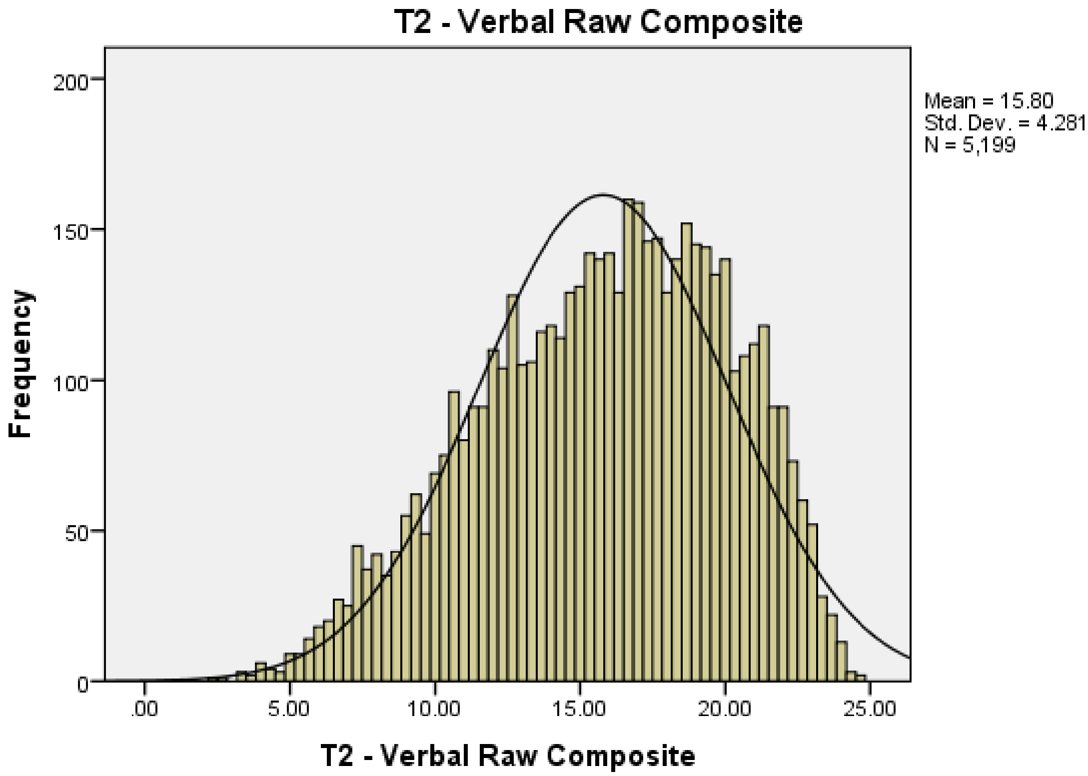

Figure 7. AFOQT Form T2 Verbal raw composite score distribution.

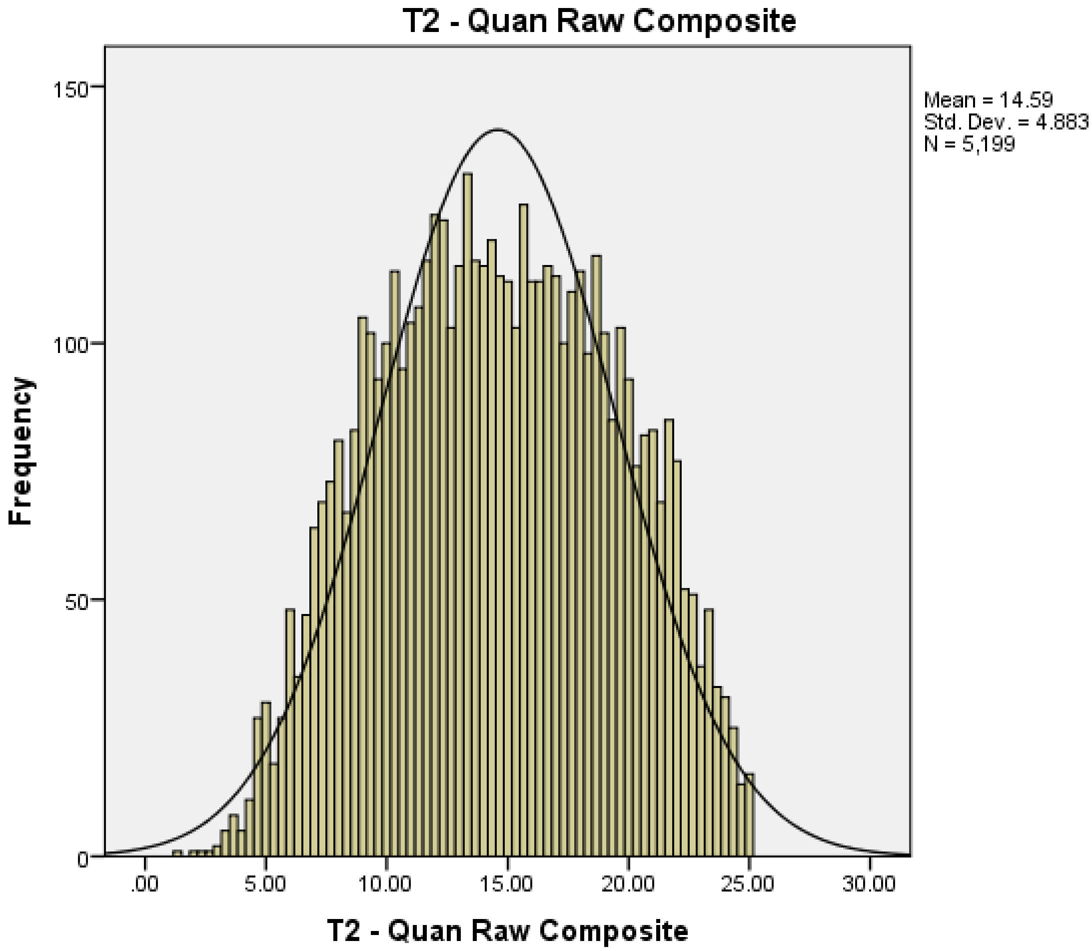

Figure 8. AFOQT Form T2 Quantitative raw composite score distribution.

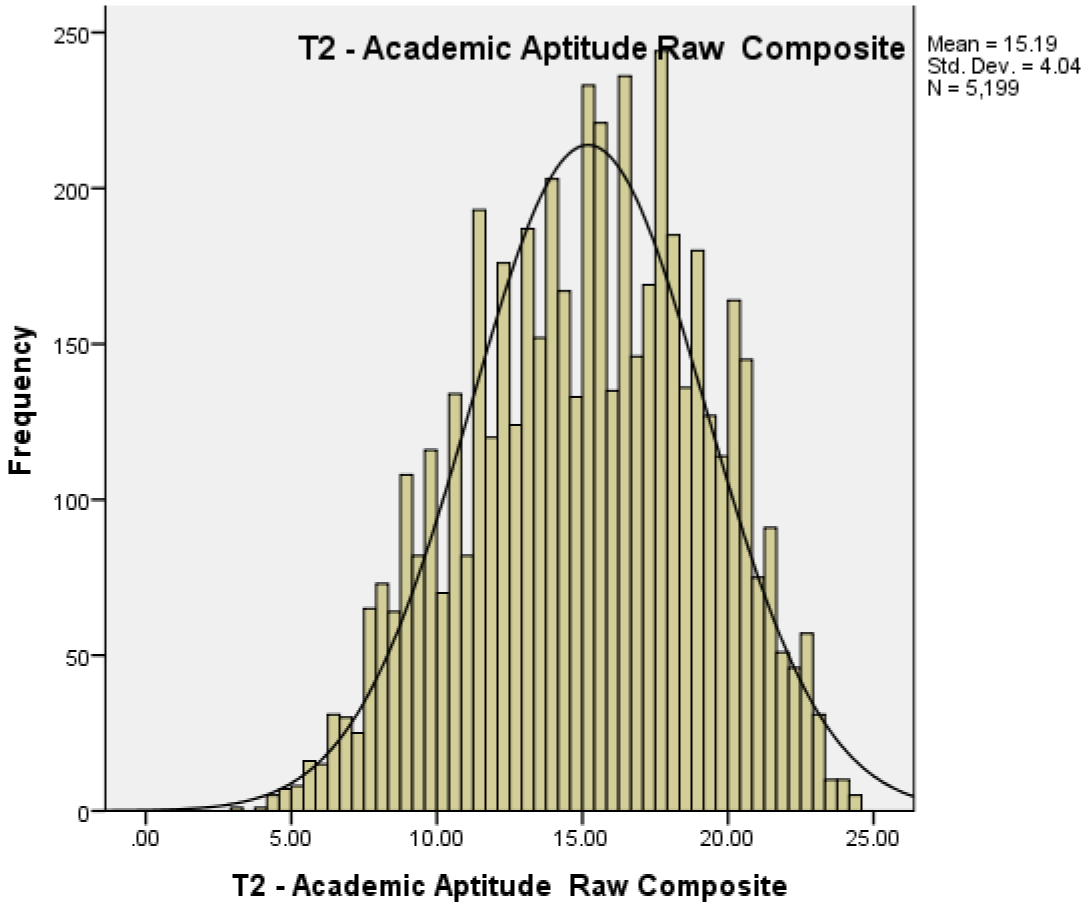

Figure 9. AFOQT Form T2 Academic Aptitude raw composite score distribution.

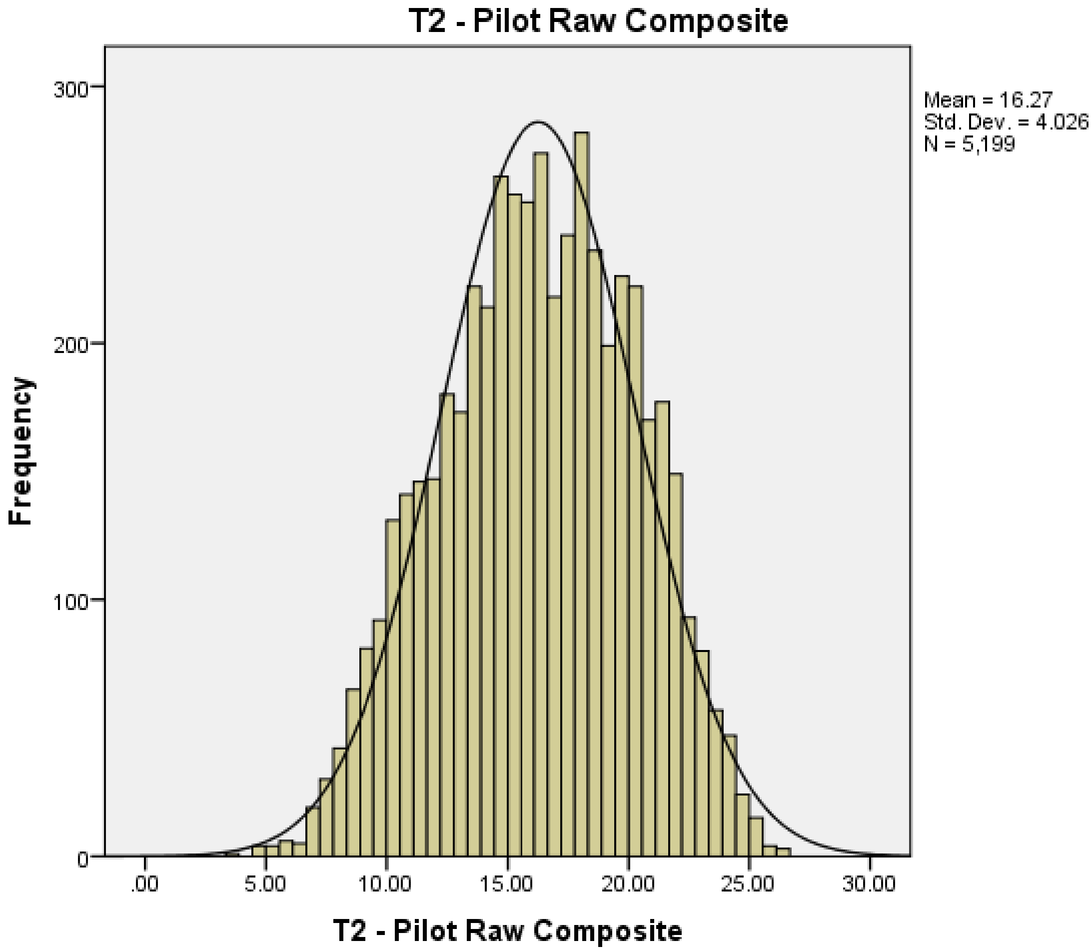

Figure 10. AFOQT Form T2 Pilot raw composite score distribution.

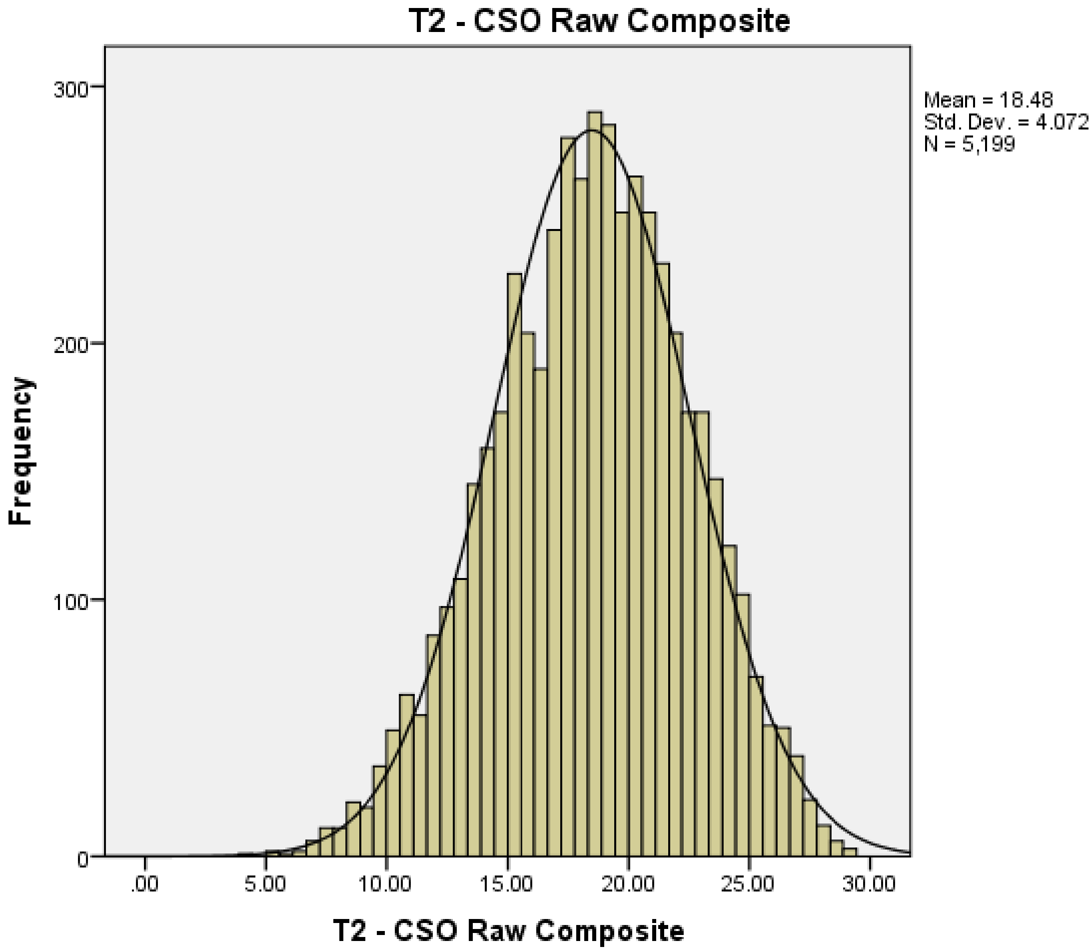

Figure 11. AFOQT Form T2 Combat Systems Officer (CSO) raw composite score distribution.

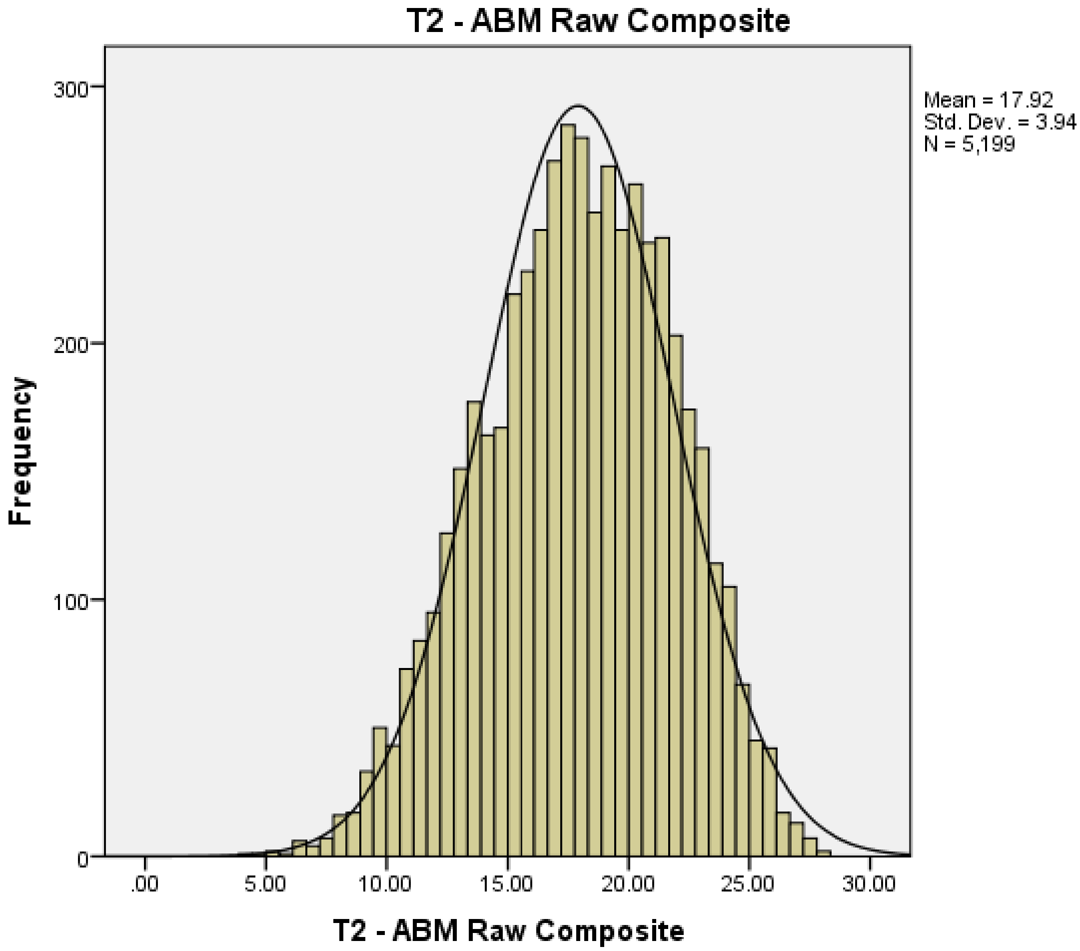

Figure 12. AFOQT Form T2 Air Battle Manager (ABM) raw composite score distribution.

Discussion

Item-Level Analyses

In general, AFOQT Forms T1 and T2 demonstrated acceptable psychometric properties. However, there were areas where improvements could be made. Examination of item-level data revealed the most difficult subtests were Physical Science, Aviation Information, and Block Counting. The high difficulty for PS and AI appear to be the result of their content. However, closer examination of the most difficult BC items indicated that difficulty level was a function of item presentation. BC includes several items where the blocks touch on their back sides out of the participant's view. However, none of the example items illustrated this condition. This can be fixed easily in the next AFOQT form by including such items in the test instructions.

Another item-level issue involves the rate of item omissions and guessing. Item omissions were low for items 1-15, but generally increased for later items. Omission rates were highest for TR (9.9%), IC (5.3%), and BC (6.9%). The omission rates in the last block of 5 items for these tests were: TR (33.2%), IC (16.2%), and BC (23.25). Currently, the AFOQT subtests are scored number correct with no penalty for guessing. It appears that some examinees are not aware that they will not be penalized for guessing, otherwise the omission rate should be low for all subtests. It is therefore recommended that instructions on guessing be further emphasized in the written (e.g., through bold text) and spoken instructions. In addition, it is recommended that policy makers evaluate the current scoring policy (i.e., no penalty for guessing) and decide whether to revise the policy to potentially enhance score precision, fairness, and validity. Further, it is recommended that the time limits for IC and BC be increased (TR is a speeded subtest) to reduce omissions.

Subtest-Level Analyses

Some subtests (VA, RC, PS, and AI) had internal consistency reliabilities below .80. Although higher reliabilities are desirable, this is not problematic as the US Air Force does not make personnel decisions based on subtest scores. Rather, personnel selection and classification decisions are based on composite scores of the subtests, which have high reliabilities.

Examination of subtest score distributions indicated that many of the distributions were non-normal, where the *t*-test for skewness, kurtosis, or both exceeded +/- 1.96. Six of 10 Form T1 subtests and 7 of 10 Form T2 subtests had skewness values greater than +/-1.96. With the exception of Aviation Information, when skewness was severe the distributions were negatively skewed. Eight of 10 Form T1 subtests and 9 of 10 Form T2 subtests had kurtosis values greater than -1.96. An examination of the score distributions (not provided in this report) revealed that with the exception of AI where kurtosis was large there was somewhat of a ceiling effect (not enough difficult items). From a psychometric standpoint, measurement of ability would be improved by adding more difficult items for all subtests except AI.

Latent Factor Structure Analyses

The *g*-saturation of Forms T1 (48.2%) and T2 (48.4%) as estimated from the first unrotated principle component were very similar to that reported for Form S (47%) by Drasgow et al. (2010). Results for model fit were mixed. Consistent with previous forms, a single factor model demonstrated poor fit. Model fit improved with the addition of lower-order factors for verbal, math, spatial, aviation, and perceptual speed. However, while values for RMSEA, SRMR, and Critical N were acceptable, those for the CFI, GFI, AGFI, and NNFI were marginal. The reasons for the somewhat lower fit compared with previous forms are not clear, but may be due to changes in content from previous forms and/or non-normality of the subtest and parcel score distributions.

Composite-Level Analyses

Composite level analyses focused on the shape of the raw score distributions. Results were consistent with those for the subtests. All composites with the exception of Quantitative were significantly negatively skewed. All had significant negative values for kurtosis. In general, the shapes for the AFOQT composite score distributions could be improved if some difficult items were added. Improving distributional shape and discriminability is more important for the aviation-related composites (Pilot, CSO, and ABM) than for the Verbal, Quantitative, or Academic Aptitude composites. This is because the minimum qualifying scores for the Verbal and Quantitative composites, which are used only for officer commissioning programs are relatively low (Verbal ≥15 and Quantitative ≥ 10). Discriminability of applicants is most important around the minimum qualifying score. Adding difficult verbal or math items would not improve discriminability for these composites. Aircrew training assignments (pilot, CSO, ABM, and RPA pilot) are much more competitive. Although minimum qualifying scores are relatively low for aircrew training (e.g., Pilot ≥ 25 for pilot training), in practice the mean Pilot composite score for those accepted into pilot training is about 80. Therefore, it is more important to improve discriminability for the aviation-related composites in the high aptitude range (i.e., greater than 70). To do so, additional difficult items are needed for some subtests that contribute (MK, TR, IC, and BC) to the aircrew-related composites (Pilot, CSO, and ABM).

References

Arth, T.O. (1986). *Validation of the AFOQT for non-rated officers,* AFHRL-TP-85-50. Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Arth, T. O., & Skinner, J. (1986). *Aptitude selection for Air Force officer non-aircrew jobs.* Paper presented at the annual meeting of the Military Testing Association, Mystic, CT.

Benson, J., & Fleishman, J. A. (1994). The robustness of maximum likelihood and distribution-free estimators to non-normality in confirmatory factor analysis. *Quality & Quantity, 28,* 117-136. Retrieved 7 July 2016 from: http://link.springer.com/article/10.1007/BF01102757#page-2

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238-246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software, Inc.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Lang (Eds.), *Testing structural equation models.* Newbury Park, CA: Sage.

Carretta. T. R. (2008). *Predictive validity of the Air Force Officer Qualifying Test for USAF air battle manager training performance,* AFRL-RH-WP-TR-2009-0007. Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate, Crew Systems Interface Division, Supervisory Control Interfaces Branch.

Carretta, T. R. (2010). Air Force Officer Qualifying Test validity for non-rated officer specialties, *Military Psychology, 22,* 450-464.

Carretta, T. R. (2013). Predictive validity of pilot selection instruments for remotely piloted aircraft training outcome. *Aviation, Space, and Environmental Medicine, 84,* 47-53.

Carretta, T. R., & Ree, M. J. (1996). Factor structure of the Air Force Officer Qualifying Test: Analysis and comparison. *Military Psychology, 8,* 29-42.

Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. In B. H. Kantowitz (Series

Ed.) & P. S. Tsang & M. A. Vidulich (Vol. Eds.). *Human factors in transportation: Principles and practices of aviation psychology* (pp. 357-396). Mahwah, NJ: Erlbaum.

Drasgow, F., Nye, C. D., Carretta, T. R., & Ree, M. J. (2010). Factor structure of the Air Force Officer Qualifying Test form S: Analysis and comparison with previous forms. *Military Psychology, 22,* 68-85.

Finegold, L., & Rogers, D. (1985). *Relationship between Air Force Officer Qualifying Test scores and success in air weapons controller training,* AFHRL-TR-85-13.Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Hartke, D. D., & Short. L. O. (1988). *Validity of the academic aptitude composite of the Air Force Officer Qualifying Test (AFOQT),* AFHRL-TP-87-61. Brooks AFB, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to

Hu, L. T.; & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Jörescog, K. G., & Sörbom, D. (1993). *LISREL 8 user's reference guide.* Chicago: science Software International.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Issues, concepts, and applications* (pp. 56-75). Newbury Park, CA: Sage.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*, 845-849.

Olsson, U. H., Foss, T., Troye, S. V., Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling, 7,* 557-565.

Ree, M. J., & Earles, J. A. (1991). The stability of convergent estimates of *g*. *Intelligence, 15,* 271-278.

Roberts, H. E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in officer training school selection decisions. *Military Psychology, 8,* 95-113.

Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*, 23-74.

Skinner, J., & Ree, M. J. (1987). *Air Force Officer Qualifying Test (AFOQT): Item and factor analysis of Form O*, AFHRL-TR-86-68. Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology, 42,* 233-239.

United States Air Force (2014). *Officer training school (OTS) and enlisted commissioning programs (ECP), Air Force Instruction 36-2013.* Washington, DC: Department of the Air Force.

List of Symbols, Abbreviations, and Acronyms

ABM	Air Battle Manager
AFOQT	Air Force Officer Qualifying Test
AFOQT T1	Air Force Officer Qualifying Test, Form T1
AFOQT T2	Air Force Officer Qualifying Test, Form T2
AGFI	Adjusted Goodness of Fit Index
AI	Air Force Officer Qualifying Test Aviation Information subtest
AR	Air Force Officer Qualifying Test Arithmetic Reasoning subtest
BC	Air Force Officer Qualifying Test Block Counting subtest
CFA	Confirmatory factor analysis
CFI	Comparative Fit Index
CSO	Combat Systems Officer
EM	Air Force Officer Qualifying Test Electrical Maze subtest
g	General mental ability factor
GFI	Goodness of Fit Index
GLS	Generalized Least Squares
GS	Air Force Officer Qualifying Test General Science subtest
IC	Air Force Officer Qualifying Test Instrument Comprehension subtest
Kurt	Kurtosis
Kurt SE	Kurtosis standard error
$\leq$	Less than or equal to
Max.	Maximum
Min.	Minimum

MK	Air Force Officer Qualifying Test Math Knowledge subtest
ML	Maximum Likelihood
N	Sample size
NNFI	Non-Normed Fit Index
OTS	Officer Training School
%	Percent
p	Probability level
PS	Air Force Officer Qualifying Test Physical Science subtest
RC	Air Force Officer Qualifying Test Reading Comprehension subtest
RMSEA	Root Mean Square Error of Approximation
ROTC	Reserve Officer Training Corps
SD	Standard deviation
Skew	Skewness
Skew SE	Skewness standard error
SRMR	Standardized Root Mean Square Residual
t	*t*-test
TR	Air Force Officer Qualifying Test Table Reading subtest
USAF	United States Air Force
VA	Air Force Officer Qualifying Test Verbal Analogies subtest
WK	Air Force Officer Qualifying Test Word Knowledge subtest
WLS	Weighted Least Squares